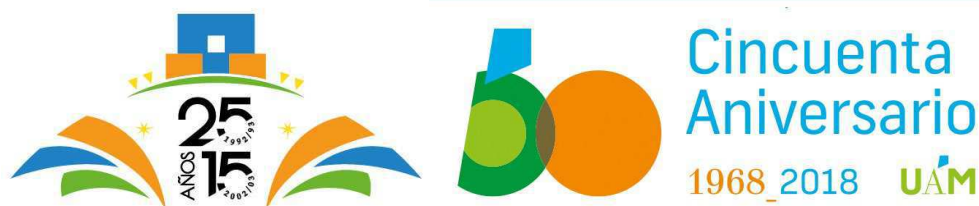


Universidad Autónoma de Madrid

Escuela politécnica superior



Máster en Bioinformática y Biología Computacional

Trabajo Fin de Máster

Desarrollo de nuevas metodologías para el análisis global de modificaciones post-traduccionales en muestras biológicas

Autor: Andrea Laguillo Gómez

Tutor: Enrique Calvo Alcocer

Ponente: Gonzalo Gómez López

Febrero 2019

ÍNDICE

1. Introducción	página 2
1.1. Importancia clínica de las PTMs	
1.2. Caracterización de PTMs: proteómica post-traducciona	
2. Objetivos	página 5
3. Materiales y Métodos	página 6
4. Resultados y Discusión	página 7
4.1. Tiempos de búsqueda	
4.2. Comparación de los valores de XCorr obtenidos	
4.3. Efecto del parámetro de tolerancia en los valores de XCorr	
4.4. Distribución de frecuencias de los picos de Δmasa observados	
4.5. Validación del incremento de XCorr mediante Vseq	
5. Conclusiones	página 17
5.1. Objetivos Futuros	
6. Bibliografía	página 19

1. Introducción

Las modificaciones post-traduccionales (PTMs) son cambios químicos en las proteínas que ocurren posteriormente a su síntesis e incrementan en gran medida la diversidad funcional de los proteomas. Es sabido que las PTMs regulan la actividad, localización e interacción de las proteínas. ^[1]

1.1. Importancia clínica de las PTMs

En muchas ocasiones, las PTMs son los principales reguladores de rutas de señalización, y estudios recientes indican que los cambios que introducen en las proteínas también podrían ser la causa o manifestación de múltiples enfermedades.

Un estudio de caracterización de PTMs específicas de tejido en ratón, realizado en este laboratorio, muestra que la existencia de heteroplasma entre variantes no patológicas de DNA mitocondrial induce una serie de modificaciones oxidativas en el corazón, que afectan predominantemente a las proteínas del sistema de fosforilación oxidativa (OXPHOS), lo cual establece una causa para las patologías asociadas a la heteroplasma. ^[2]

Otro ejemplo de la importancia clínica de las PTMs es su posible utilidad como biomarcadores para algunas enfermedades, potencialmente permitiendo una detección temprana cuando aún se encuentran en estado subclínico. Múltiples estudios indican que las PTMs y las modificaciones estructurales juegan un papel importante en el desarrollo de la aterosclerosis. La oxidación de LDL es un evento clave en la patogénesis de esta enfermedad. ^[3] Las glicosilaciones, modificaciones oxidativas y variantes de secuencia codificante también están relacionadas con el infarto agudo de miocardio, la formación de placas de ateroma y el desarrollo de aterosclerosis subclínica. ^{[4][5]} Además, algunos estudios sugieren que la citrulinación y la carbamilación podrían estar implicadas en la patogénesis de la aterosclerosis. ^[6] En conjunto, estos resultados indican que las PTMs son relevantes en el desarrollo de esta enfermedad, y potencialmente podrían ser usadas como biomarcadores.

1.2. Caracterización de PTMs: proteómica post-traduccional

Las PTMs pueden analizarse mediante espectrometría de masas shotgun y algoritmos basados en búsquedas en bases de datos. ^[7] Sin embargo, sólo una minoría de los espectros obtenidos pueden ser asignados a proteínas, y se cree que una gran proporción de los espectros no asignados provienen de péptidos que contienen variaciones de secuencia o modificaciones post-traduccionales o químicas desconocidas. ^[8]

Se han desarrollado estrategias de búsqueda abierta (búsquedas ultratolerantes, “Open Search”), en las cuales se emplean tolerancias del orden de cientos de Dalton para la masa del precursor, con bases de datos convencionales. ^[9] De esta forma se logran identificar péptidos modificados a una escala sin precedentes. Aun así, estas aproximaciones sólo son capaces de identificar la mitad de las modificaciones potencialmente detectables, ^[10] y no localizan la posición del residuo modificado. Por otro lado, existe una falta de herramientas para el análisis cuantitativo de PTMs.

Comet MS/MS ^[11] es un software escrito en C++ para la búsqueda de espectros de masas de péptidos en bases de datos de secuencias con el objetivo de identificar péptidos y proteínas presentes en muestras biológicas. Comet-PTM ^[2] es una versión de Comet MS/MS desarrollada en nuestro laboratorio de Proteómica Cardiovascular dirigido por el Prof. Jesús Vázquez, que mejora la identificación de

modificaciones post-traduccionales y localiza la posición del residuo modificado, teniendo en cuenta la variación de masa (Δmasa) que aparece en las series de fragmentación de un péptido modificado. Este método produce puntuaciones (XCorr) que son:

- Equivalentes a las de una búsqueda cerrada (“Closed Search”) convencional (es decir, con valores de tolerancia pequeños) si se emplea dicha Δmasa como modificación variable (Figura 1).
- Superiores a las de una búsqueda abierta convencional (Figura 1).

El valor de XCorr es obtenido puntuando el número de iones observados en el espectro de fragmentación frente al espectro teórico del candidato, y calculando la puntuación de correlación cruzada para todos los péptidos candidatos obtenidos de la base de datos empleada. [12]

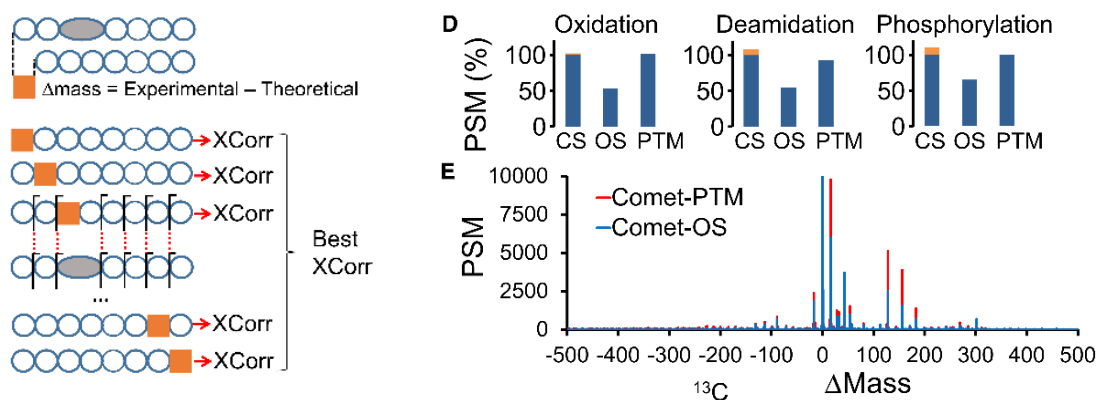


Figura 1. Comet-PTM duplica el rendimiento de los algoritmos de Open Search existentes, permitiendo por primera vez la generación de mapas exhaustivos de péptidos de forma no sesgada, incluyendo prácticamente todas las modificaciones detectables mediante búsquedas Closed Search. Así, Comet-PTM abre el camino hacia el análisis de PTMs de forma libre de hipótesis (hypothesis-free). [12]

Comet-PTM calcula la variación de masa experimental correspondiente a un espectro comparando la masa teórica según la secuencia peptídica con la masa experimental detectada por el espectrómetro de masas, de acuerdo a la siguiente fórmula:

$$\Delta\text{masa experimental} = \text{masa teórica} - \text{masa experimental}$$

Esta Δmasa experimental puede no corresponderse exactamente con la Δmasa teórica para cada modificación posible por diversos motivos, tales como:

- Inexactitud en la determinación de la masa experimental del precursor, debido a errores del analizador (espectrómetro de masas) (Figura 2).
- Error en la selección del pico monoisotópico en el espectro, que puede ocurrir cuando dicho pico no es el de mayor intensidad (Figura 2).

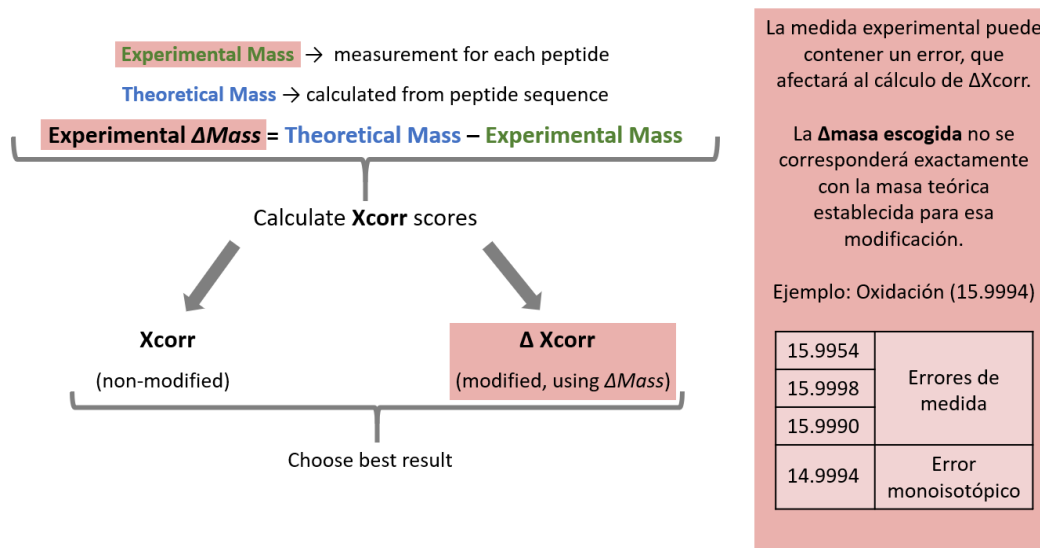


Figura 2. Algoritmo para el cálculo de Xcorr en Comet-PTM.

Cuando la Δ masa experimental presenta este tipo de problemas, la identificación de la modificación presente en el péptido en cuestión puede ser incorrecta. En este caso, proponemos que la sustitución de esta Δ masa experimental por la Δ masa teórica exacta para esa modificación podría mejorar la identificación de la modificación, reflejada en un aumento de la puntuación (Xcorr) asignada a ese candidato. Para lograr este objetivo, nos planteamos modificar Comet-PTM para añadir funcionalidades que permitan mapear la Δ masa experimental en un listado de masas teóricas correspondientes a modificaciones conocidas dentro de un intervalo de masas previamente fijado (Figura 3).

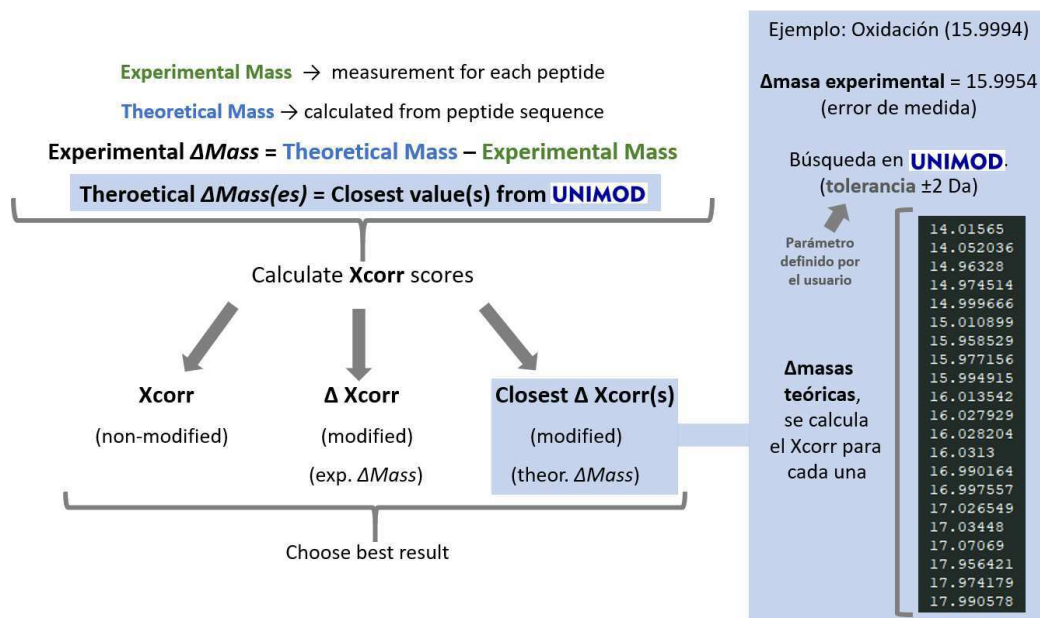


Figura 3. Modificaciones propuestas para mejorar la determinación de los mejores valores de Δ masa en Comet-PTM. El parámetro de tolerancia para definir el rango de masas teóricas a considerar puede variar para adecuarse al tipo de experimento: en el caso de experimentos DDA (Data-Dependent Acquisition) es suficiente con valores bajos de tolerancia, mientras que para experimentos DIA (Data-Independent Acquisition) se requieren valores altos de tolerancia, ya que no se dispone de la masa experimental del ión, sino un rango de posibilidades, por ello la Δ mass podrá tener alto error.

2. Objetivos

Con el objetivo de mejorar la identificación de PTMs en las búsquedas con Comet-PTM, planteamos realizar modificaciones a dicho software que permitan:

1. Mapear la Δ masa experimental obtenida para un espectro con una lista de masas de modificaciones teóricas establecidas (por ejemplo, las recogidas en la base de datos UNIMOD^[13]) e identificar los candidatos más probables, por cercanía a la Δ masa experimental y dentro de un rango de tolerancia definido por el usuario.
2. Calcular la puntuación (XCorr) para cada una de las Δ masas teóricas identificadas en el paso anterior.
3. Seleccionar aquella Δ masa teórica que haya alcanzado la mejor puntuación y añadir dicha información al output de Comet-PTM para cada péptido.

Mediante esta estimación de Δ masa de forma semi-supervisada, se pretende conseguir una mejor interpretación de los datos obtenidos por espectrometría de masas, así como crear una estrategia de búsqueda compatible con experimentos DIA (Data Independent Acquisition), en los cuales los errores de Δ masa van a ser mayores ya que no se dispone de una masa precisa para los precursores fragmentados.

3. Materiales y Métodos

DESARROLLO DEL SOFTWARE Y ANÁLISIS:

Las modificaciones para el cálculo de XCorr con Δ masas teóricas se realizaron sobre el código original de Comet-PTM (C++). Comet-PTM es software libre y el código fuente se encuentra disponible para su descarga. ^[14] Se realizaron búsquedas con la versión original de Comet-PTM y la versión modificada, denominada RECOM (*Refined Comet*), empleando el clúster disponible en el CNIC, en el cual cada nodo está compuesto por 2 x Intel Xeon E5-2695v2 a 2.40 GHz y contiene 46 hilos/124 gigabyte.

La lista de Δ masas teóricas se introduce en RECOM a partir de la lectura de un archivo de texto, proporcionado por el usuario como parámetro de búsqueda. Se introdujo una función en RECOM para extraer de dicha lista las Δ masas contenidas dentro del valor de tolerancia, para cada candidato analizado. Comet-PTM contiene una función para el cálculo del XCorr, que en RECOM se ha modificado de forma que itere sobre el conjunto de Δ masas teóricas seleccionadas, obteniendo la puntuación para cada una y devolviendo únicamente los resultados correspondientes a la Δ masa que produce una mejor puntuación. RECOM devuelve los XCorr obtenidos con la Δ masa experimental y la mejor Δ masa teórica para cada péptido, permitiendo la comparación de puntuaciones.

PARÁMETROS DE BÚSQUEDA:

Para las búsquedas se empleó un dataset procedente del análisis por MS de una muestra de plasma humano marcado con Tandem Mass Tags (TMT 6plex) de un proyecto en curso de colaboración del laboratorio. Se empleó una base de datos concatenada target/decoy que contiene las proteínas humanas recogidas en SwissProt (versión UniProt_2018_07). ^[15]

Se establecieron los siguientes parámetros de búsqueda para la ejecución de Comet-PTM y RECOM:

- Tolerancia para la masa del ión precursor: 500 Da
- Tolerancia para la masa de los fragmentos: 0.01 Da
- Método de activación: HCD (Higher-energy Collisional Dissociation)
- Rango de masas MH+ que analizar: 400.0 - 6000.0 Da
- Tolerancia para el valor de Δ masa experimental: 500 Da
- Tolerancia para los valores de Δ masa teóricos: 1Da (para la comparativa de puntuaciones obtenidas entre 1Da y 2Da) y 2Da (para el resto de resultados).

La identificación de las Δ masas más cercanas se realizó a partir del listado de modificaciones descritas en la base de datos UNIMOD, incluyendo de forma manual los valores correspondientes a la presencia de 1 o 2 átomos de Carbono-13 (1.00335483 Da y 2.00670966 Da, respectivamente).

VSEQ:

Vseq es una aplicación de R desarrollada en nuestro laboratorio ^[18] para el cálculo de E-scores y la elaboración de gráficas de control de calidad para controlar la especificidad de la identificación de péptidos. Se ha empleado esta aplicación para comprobar las mejoras de las puntuaciones en los iones identificados mediante el RECOM. Vseq es software libre y el código de R se encuentra disponible para su descarga. ^[16]

4. Resultados y Discusión

4.1. Tiempos de búsqueda

Se compararon los tiempos de búsqueda para un dataset de 10 000 scans, empleando los mismos parámetros de búsqueda para la versión original de Comet, Comet-PTM y RECOM con los cambios propuestos para el cálculo de XCorr con Δ masas teóricas (Figura 4). En el caso de RECOM contamos con un nuevo parámetro, la tolerancia para los valores de Δ masa teóricos. Se realizó la misma búsqueda con valores de ± 1 Da y ± 2 Da para este parámetro.

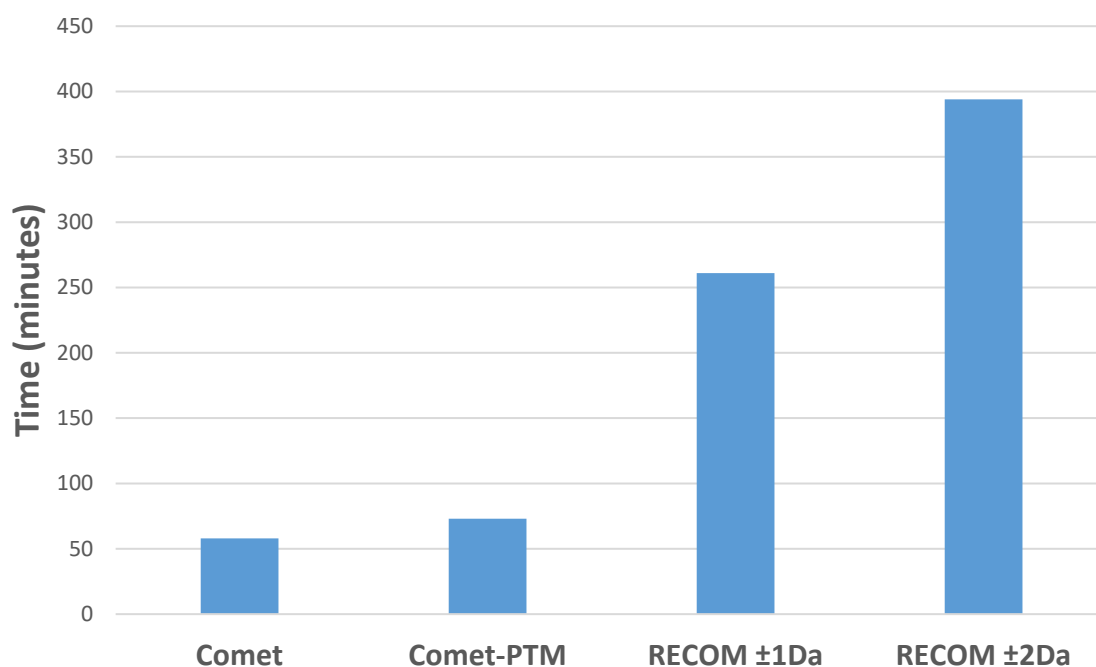


Figura 4. Tiempos de búsqueda para las distintas versiones de Comet.

Observamos que RECOM es aproximadamente 4.5 veces más lento que Comet, y 3.5 veces más lento que Comet-PTM, empleando un valor de ± 1 Da de tolerancia para los valores de Δ masa teóricos. Para un valor de ± 2 Da, RECOM es aproximadamente 6.8 veces más lento que Comet, y 5.4 veces más lento que Comet-PTM. Este aumento de la duración de las búsquedas es debido a la necesidad de repetir los cálculos de XCorr con las múltiples Δ masas teóricas que encajan en la ventana previamente definida. Al aumentar el valor de tolerancia, aumenta el número de Δ masas que será necesario evaluar, y por lo tanto el tiempo de búsqueda. Consideramos que una tolerancia de ± 2 Da debería ser suficiente para la mayoría de situaciones, ya que incluye prácticamente cualquier error de medida experimental, incluyendo los saltos monoisotópicos debido a la presencia de Carbono-13.

4.2. Comparación de los valores de XCorr obtenidos

En primer lugar, comparamos los valores de XCorr obtenidos para la misma búsqueda realizada con Comet-PTM frente a RECOM, con una tolerancia de ± 2 Da para las Δ masas teóricas (Figura 5). En esta gráfica, los puntos situados a lo largo de la diagonal se corresponden a identificaciones cuyo XCorr es idéntico para ambas versiones de Comet-PTM, es decir, ninguna Δ masa teórica produce un XCorr superior a la Δ masa experimental establecida originalmente por Comet. Aquellos puntos situados por encima de la diagonal representan identificaciones para las cuales el nuevo cálculo de XCorr con alguna Δ masa teórica dentro del rango establecido ha generado una puntuación superior a la obtenida con la Δ masa experimental.

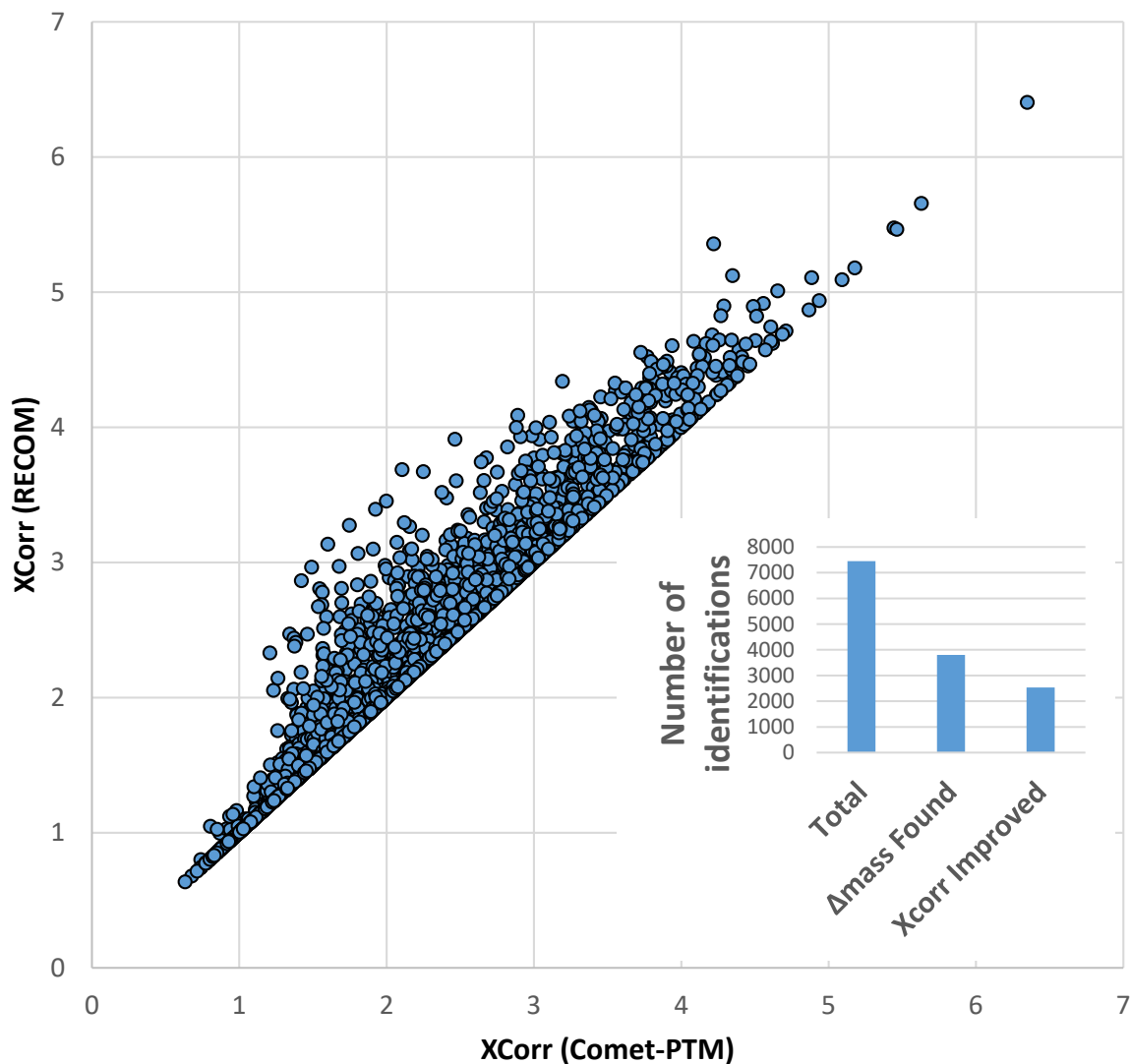


Figura 5. Comparación de los valores de XCorr obtenidos con Comet-PTM y RECOM. Se representan únicamente los valores de XCorr para el primer candidato seleccionado para cada scan (Comet-PTM devuelve por defecto los 5 mejores candidatos para cada scan). El gráfico de barras muestra el total de identificaciones realizadas, el número de identificaciones para las cuales se encontró al menos una Δ masa teórica (51.00%) y la proporción de éstas para las cuales el XCorr mejoró empleando la Δ masa teórica (66.69%, o 34.01% del total).

Considerando únicamente el primer candidato (el de mayor XCorr obtenido en la búsqueda) para cada scan analizado, de un total de 7452 identificaciones, encontramos masas teóricas en el rango especificado (± 2 Da) para 3801 (51.00%). Entre estas últimas, en 2535 scans se obtuvo una mejora del XCorr, lo cual supone un 66.69% de los casos recalculados y un 34.01% del total de scans.

4.3. Efecto del parámetro de tolerancia en los valores de XCorr

En la *Figura 6* comparamos los valores de XCorr para los primeros candidatos de cada scan para la misma búsqueda realizada con tolerancias de ± 1 Da y ± 2 Da, comprobamos que solo se incrementa el porcentaje de XCorr mejorados en un 11% de los casos cuando usamos la ventana mayor (± 2 Da). Este porcentaje podría ser considerado bajo teniendo en cuenta el notable incremento del tiempo necesario para el análisis. Sin embargo, la ventana ampliada de 2 Da sería necesaria para casos concretos en los que se haya hecho un análisis basado en DIA, quedando la ventana de 1 Da para los experimentos DDA.

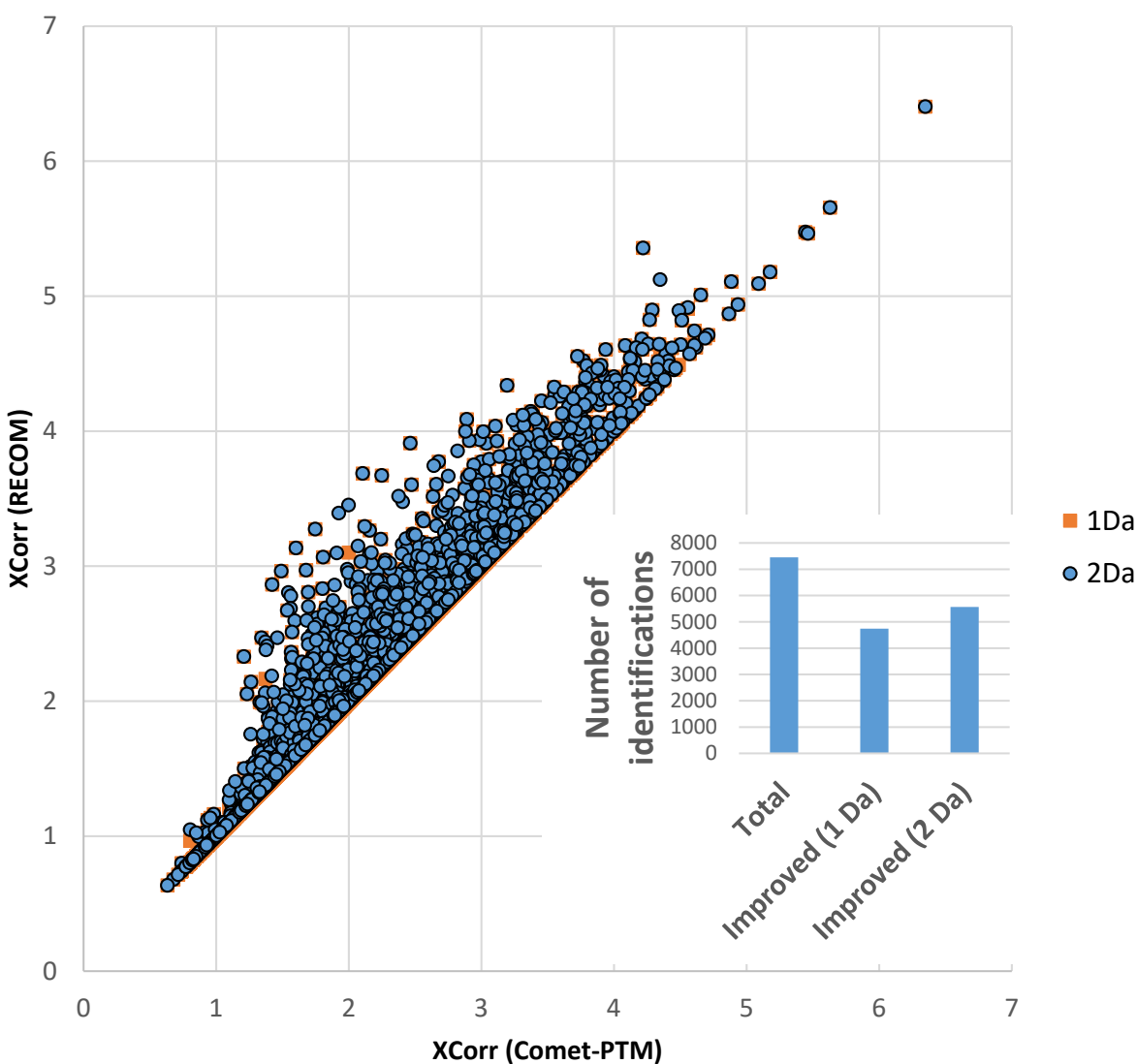


Figura 6. Comparación de los valores de XCorr obtenidos con Comet-PTM y RECOM, con valores de tolerancia para las Δ masas teóricas de ± 1 Da y ± 2 Da.

4.4. Distribución de frecuencias de los picos de Δ masa observados

Al comparar los histogramas de Δ masas encontradas para los primeros candidatos en búsquedas realizadas con Comet-PTM y RECOM comprobamos que la versión modificada de Comet-PTM logra concentrar de una manera muy clara los picos de Δ masas en los valores teóricamente correctos (tal y como se encuentran definidos en UNIMOD), eliminando gran parte de los errores experimentales y saltos monoisotópicos que han sido sustituidos por la Δ masa teórica siempre que produzca una mejor puntuación, tal como hemos descrito en el apartado 4.2. (Figuras 7-11) Los histogramas presentados a continuación se han representado de forma que 1 bin equivale a 0.002 Da.

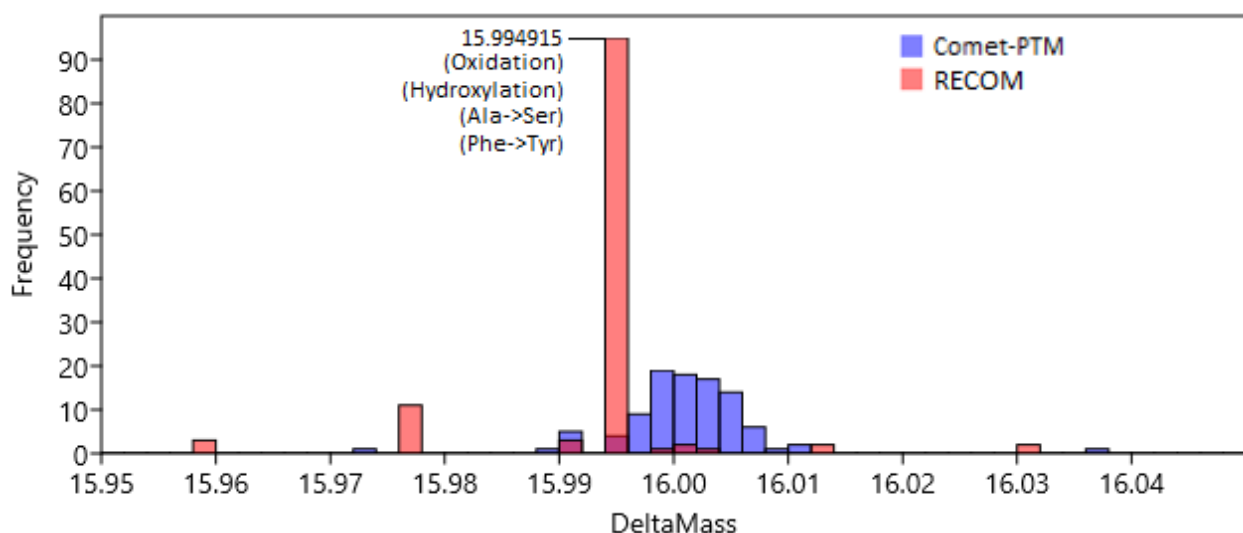


Figura 7. El pico correspondiente a oxidaciones (Δ masa teórica de 15.994915 Da) contiene valores experimentales muy dispersos, y una gran mayoría producen un mejor XCorr al sustituir el valor experimental por el teórico.

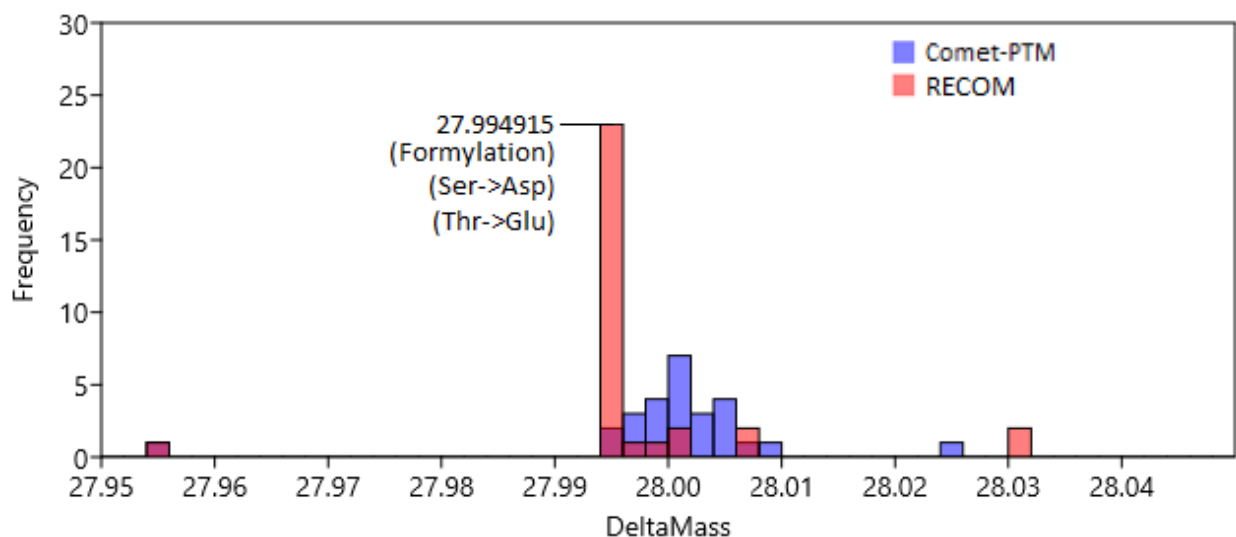


Figura 8. Para esta búsqueda, Comet-PTM encuentra múltiples Δ masas experimentales en torno a 28 Da, que RECOM identifica en su mayoría como formilaciones (Δ masa teórica de 15.994915 Da), una modificación artefactual introducida en los péptidos como consecuencia del empleo de ácido fórmico en la preparación de las muestras para su análisis por espectrometría de masas. Esta Δ masa teórica también podría corresponder a una sustitución Ser->Asp o Thr->Glu, para los péptidos que contengan dichos aminoácidos en su secuencia.

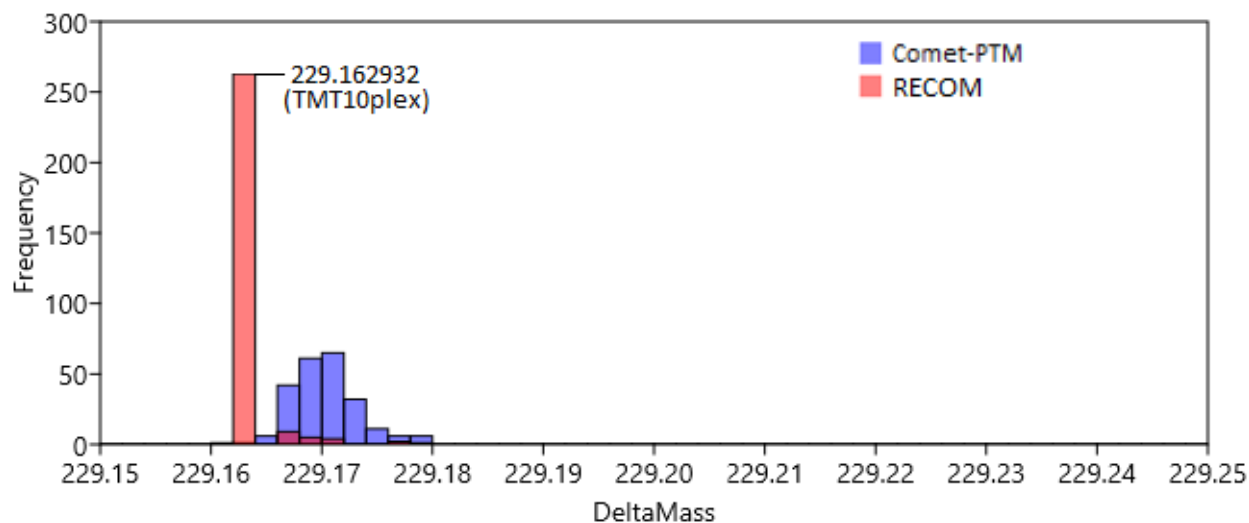


Figura 9. Comet-PTM también encuentra múltiples Δ masas experimentales en torno a 229 Da, que RECOM identifica prácticamente en su totalidad como TMT 10 plex, el marcaje isobárico empleado en este experimento para su análisis por espectrometría de masas. Este resultado confirma la correcta presencia del marcaje en la muestra.

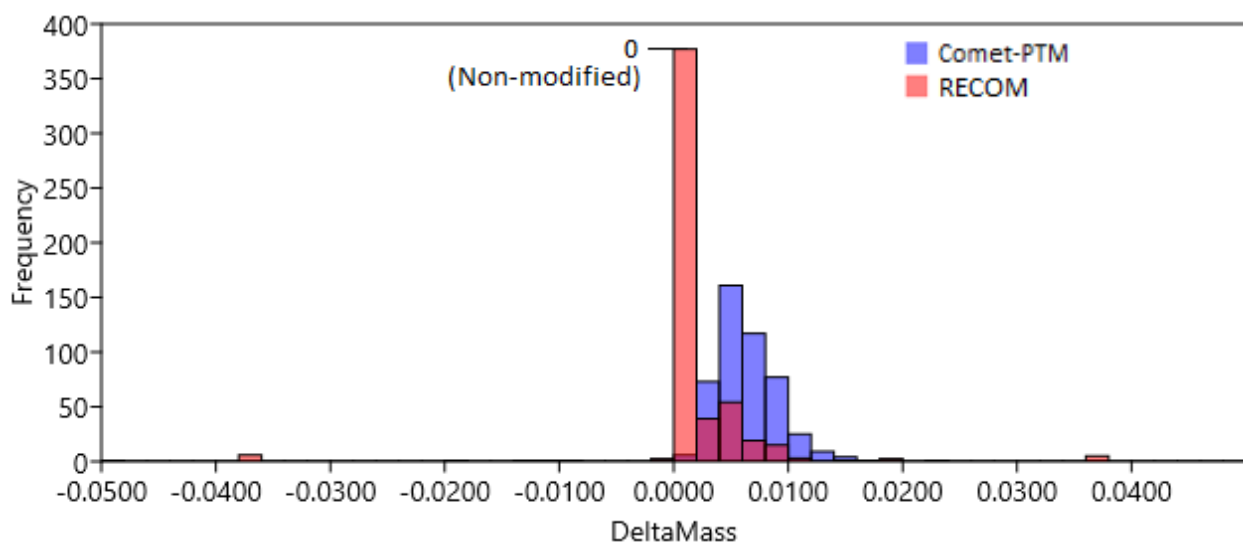


Figura 10. Observamos que múltiples Δ masas experimentales cercanas a 0 Da han sido corregidas por RECOM, confirmando que son mayoritariamente péptidos no modificados que han sufrido errores de medición. Se trata simplemente de una comprobación, ya que Comet-PTM incluye por defecto el cálculo de XCorr no modificado para cada péptido candidato.

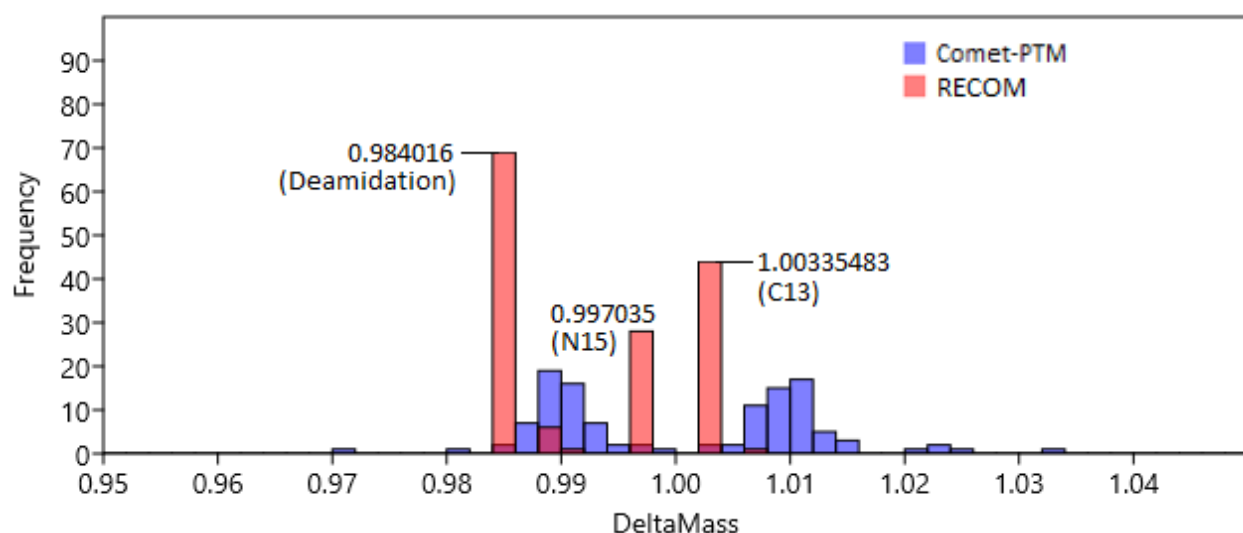


Figura 11. Al añadir a la lista de Δ masas los valores correspondientes a la presencia de 1 Carbono-13, observamos que se logran corregir múltiples Δ masas teóricas cercanas a ese valor, que previamente no se asociaban a ninguna modificación teórica. Cabe destacar que en este caso los resultados parecen indicar, además, la presencia de Nitrógeno-15 en la muestra a niveles comparables a los de Carbono-13. Esto no se corresponde a lo esperado, ya que el Nitrógeno-15 es muy poco abundante en la naturaleza (0.3%) comparado con el Carbono-13 (1.1%). Este resultado podría explicarse si esas Δ masas realmente corresponden a alguna Δ masa teórica no presente en UNIMOD y muy cercana al valor del N15, que produciría un aumento de XCorr simplemente por ser más cercano a la Δ masa teórica correcta que la Δ masa experimental observada. Por último, muchos valores de Δ masa cercanos a 1 Da pasan a estar asociados a la Δ masa teórica correspondiente a una deamidación.

4.5. Validación del incremento de XCorr mediante Vseq

Se realizó un análisis mediante Vseq de algunos de los scans para los cuales RECOM detecta un incremento de XCorr (Figuras 13-17). El objetivo es valorar cómo cambia la cobertura y la calidad de las fragmentaciones en cada uno de los scans cuyo XCorr ha mejorado cuando se ha calculado la Δ masa teórica. Vseq representa 3 gráficas del análisis para cada uno de los scans:

- Una distribución de intensidades de acuerdo con el error en partes por millón (ppm). Únicamente se considerará que pertenecen al scan aquellos fragmentos para los cuales el error sea inferior a 50 ppm, y sólo aquellos con error inferior a 9 ppm contribuyen al E-score (parámetro semi-cuantitativo).
- El espectro de fragmentación correspondiente a dicho scan con las series principales de fragmentación (*b* e *y*) debidamente asignadas.
- Una distribución de los iones observados para las series *b* e *y*, que muestra la cobertura de la secuencia de acuerdo con un código colorimétrico (Figura 12) que determina el error en ppm. La serie *b* de fragmentación se representa en la mitad izquierda del gráfico, en orden creciente, mientras que la serie *y* de fragmentación se representa en la mitad derecha del gráfico de forma inversa. Un péptido para el cual la cobertura fuese total produciría una representación completa en forma de "V". Los fragmentos marcados con un triángulo blanco son aquellos que portan el incremento de masa debido a la modificación.



Figura 12. Código colorimétrico según el error en ppm.

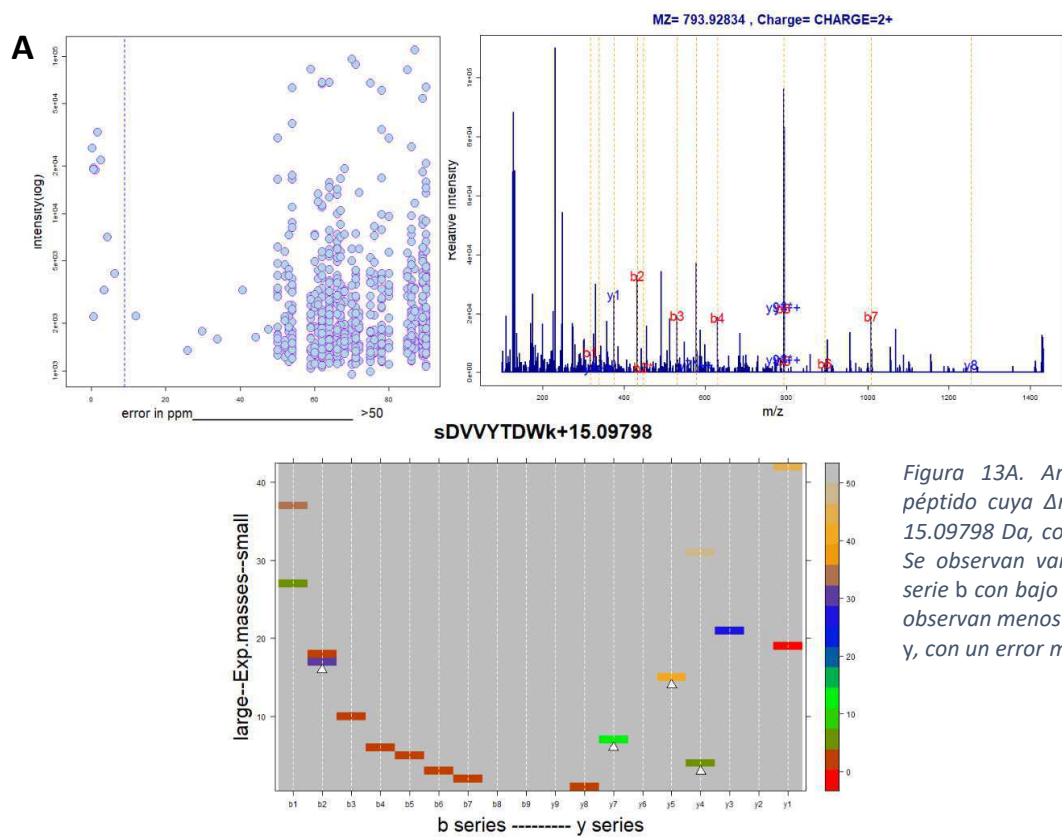


Figura 13A. Análisis Vseq para un péptido cuya Δ masa experimental es 15.09798 Da, con un XCorr de 1.4917. Se observan varios fragmentos de la serie b con bajo error, mientras que se observan menos fragmentos de la serie y, con un error mucho más alto.

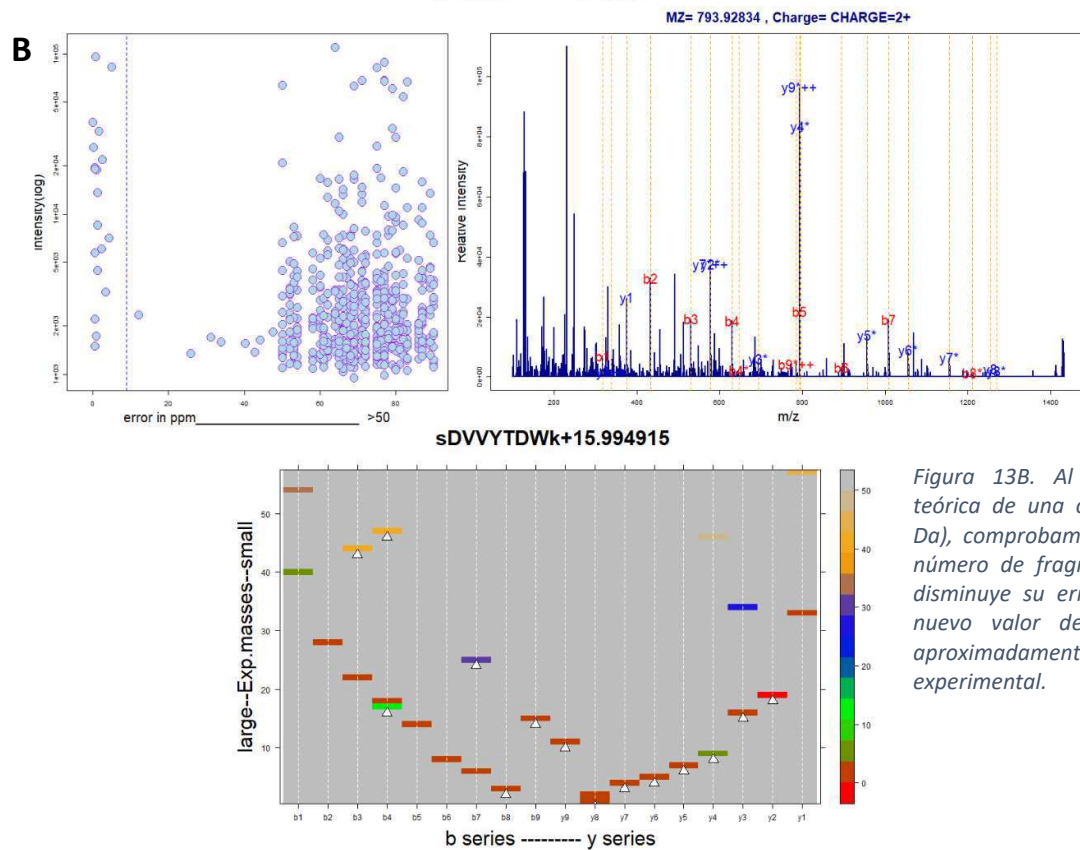


Figura 13B. Al emplear la Δ masa teórica de una oxidación (15.994915 Da), comprobamos cómo aumenta el número de fragmentos detectados y disminuye su error, obteniéndose un nuevo valor de XCorr de 2.9631, aproximadamente el doble que el valor experimental.

A

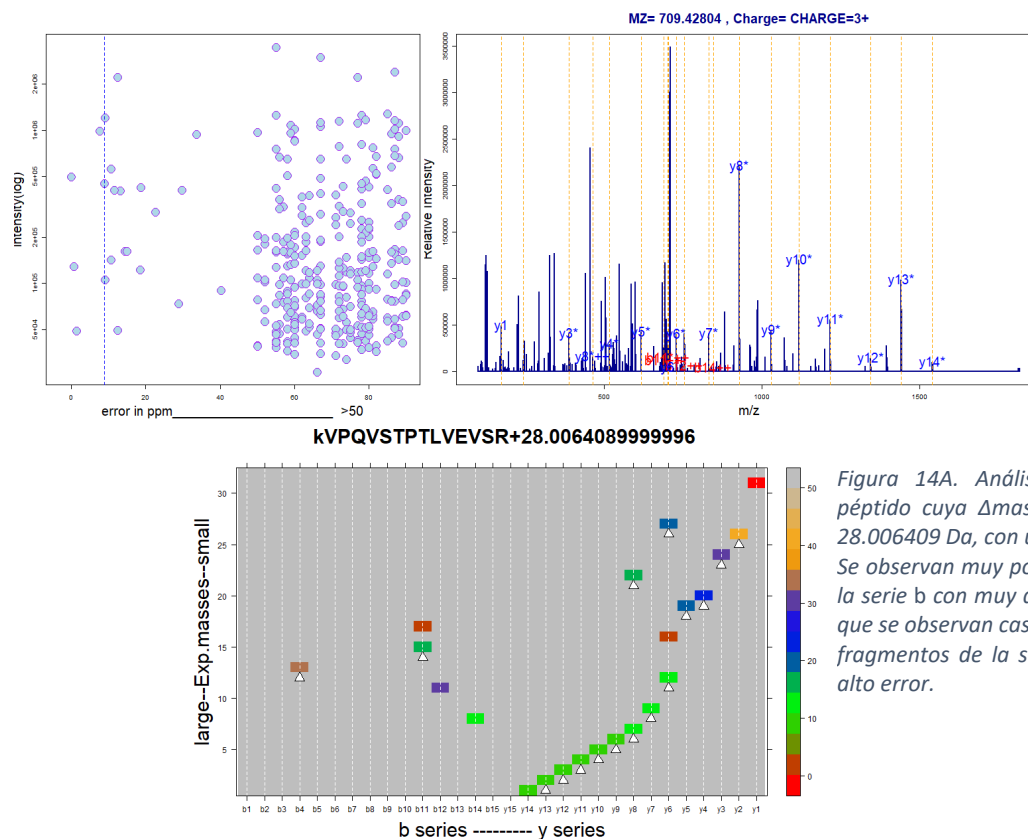


Figura 14A. Análisis Vseq para un péptido cuya Δ masa experimental es 28.006409 Da, con un XCorr de 3.1947. Se observan muy pocos fragmentos de la serie b con muy alto error, mientras que se observan casi la totalidad de los fragmentos de la serie y, también con alto error.

B

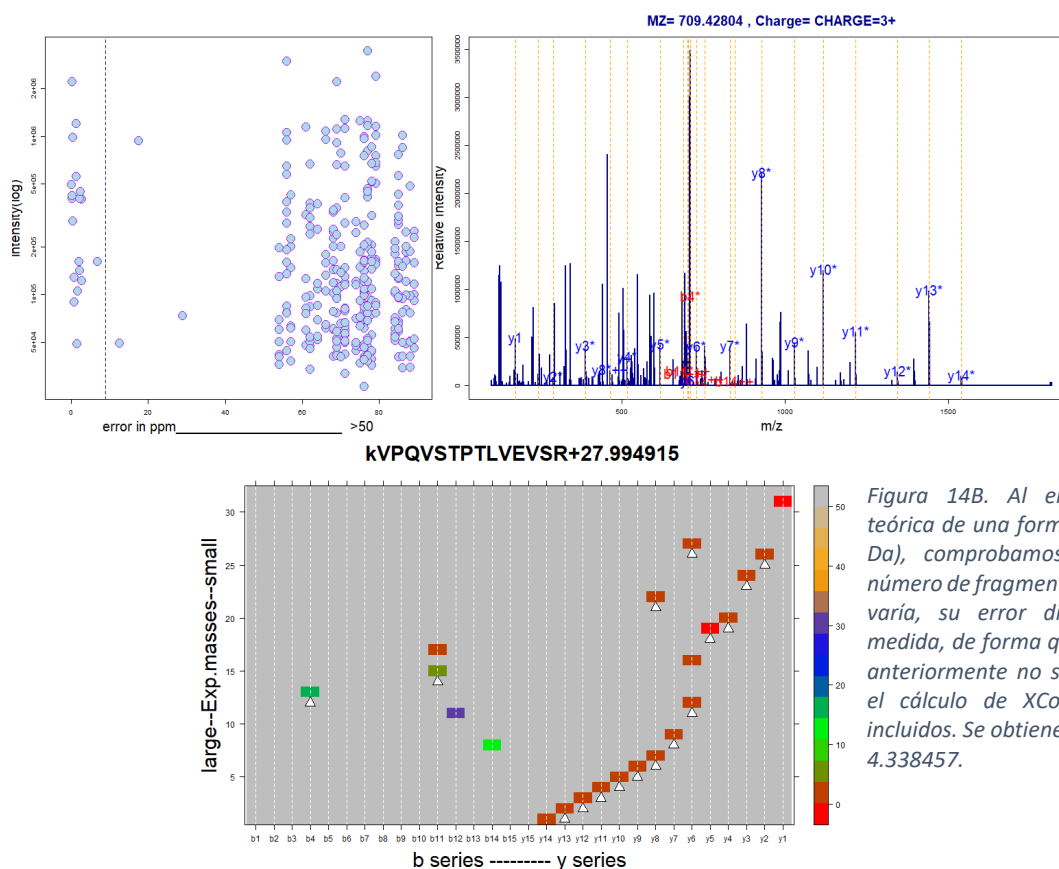


Figura 14B. Al emplear la Δ masa teórica de una formulación (27.994915 Da), comprobamos que, aunque el número de fragmentos identificados no varía, su error disminuye en gran medida, de forma que fragmentos que anteriormente no se consideraban en el cálculo de XCorr pasan a estar incluidos. Se obtiene un nuevo XCorr de 4.338457.

A

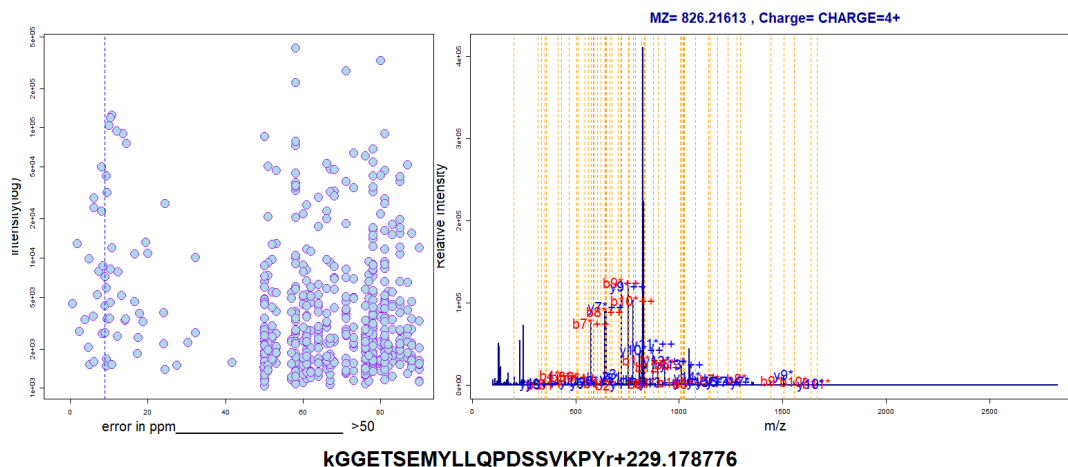


Figura 15A. Análisis Vseq para un péptido cuya Δ masa experimental es 229.178776 Da, con un XCorr de 4.2200. Se observa una gran mayoría de fragmentos de ambas series de fragmentación, pero los valores de error son altos o muy altos para la mayoría de fragmentos.

B

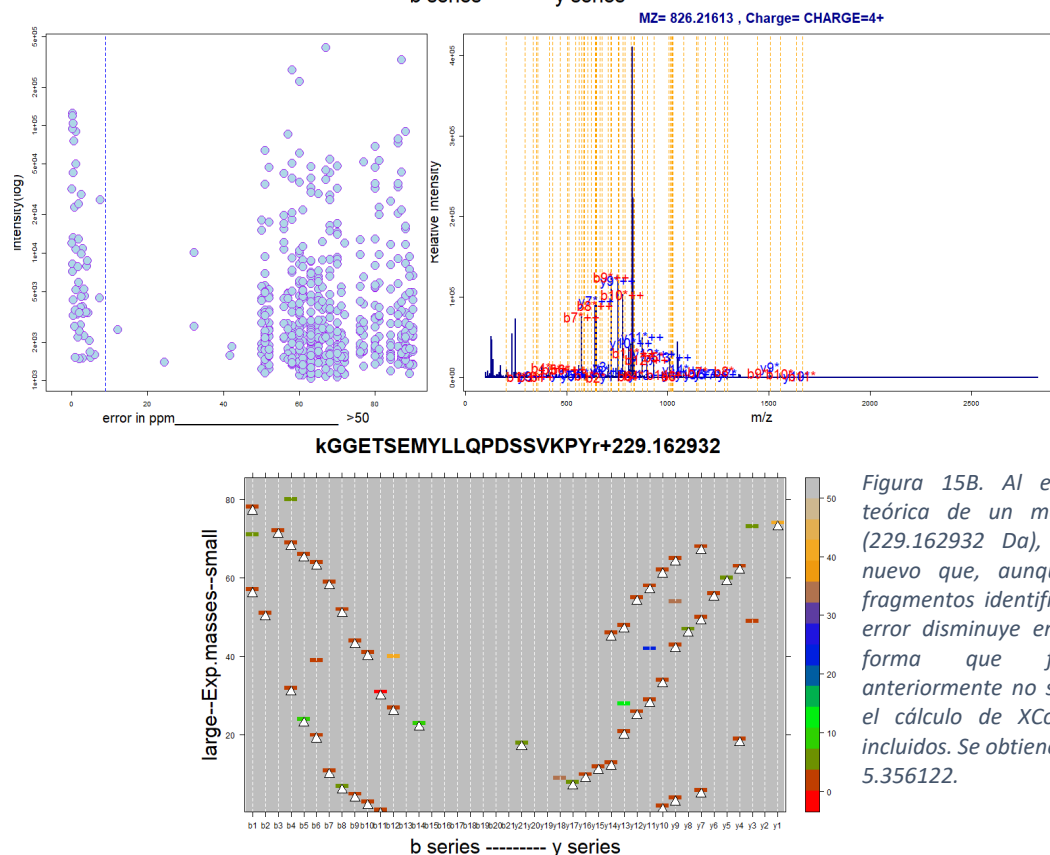
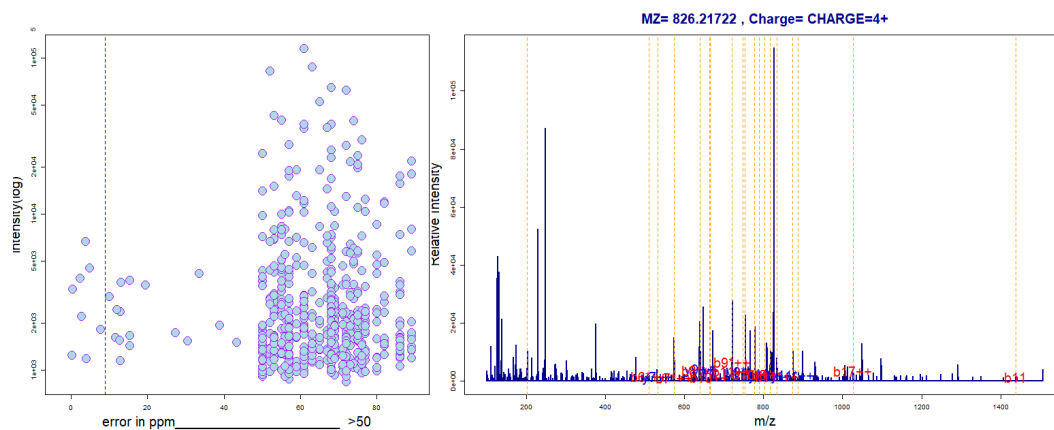


Figura 15B. Al emplear la Δ masa teórica de un marcaje TMT 6plex (229.162932 Da), comprobamos de nuevo que, aunque el número de fragmentos identificados no varía, su error disminuye en gran medida, de forma que fragmentos que anteriormente no se consideraban en el cálculo de XCorr pasan a estar incluidos. Se obtiene un nuevo XCorr de 5.356122.

A



kGGETSEMYLLQPDSSVKPYr+228.178044

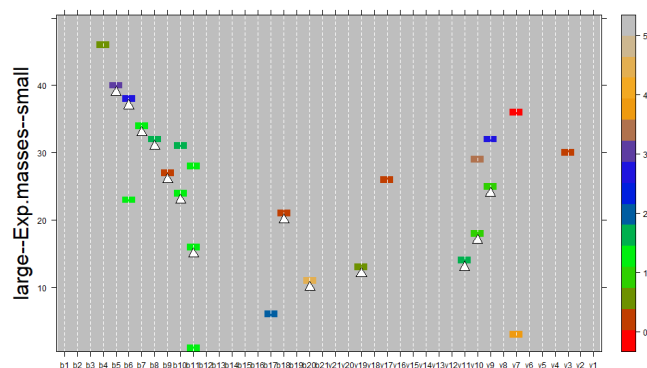
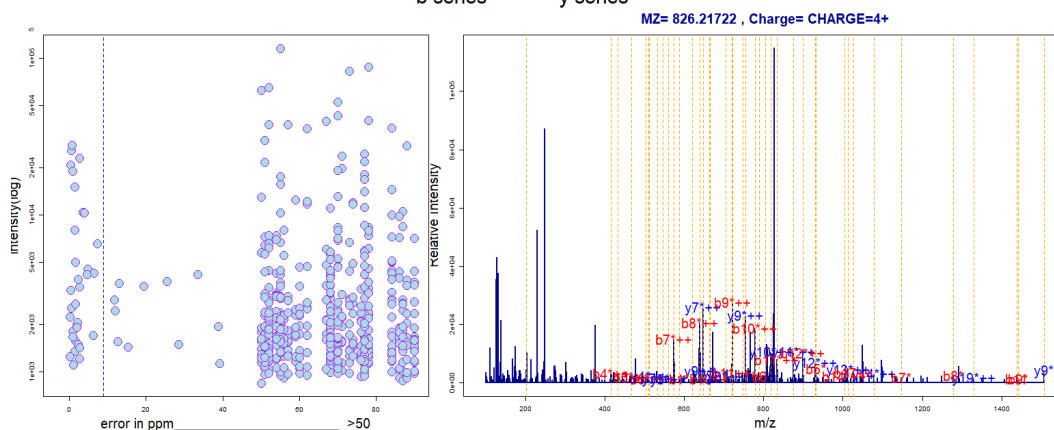


Figura 16A. Análisis Vseq para un péptido cuya Δ masa experimental es 228.178044 Da, con un XCorr de 2.1061. Se observa que ambas series de fragmentación aparecen de forma muy incompleta, con valores altos de error en la mayoría de los casos.

B



kGGETSEMYLLQPDSSVKPYr+229.162932

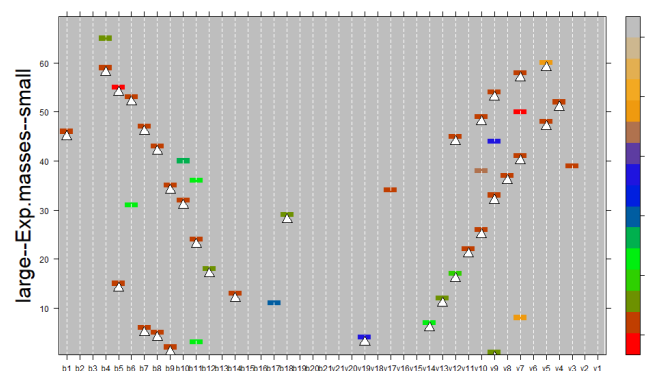


Figura 16B. Al emplear la Δ masa teórica de un marcaje TMT 6plex (229.162932 Da), comprobamos que se completan en gran medida ambas series de fragmentación, con niveles de error muy inferiores. De esta forma se obtiene un nuevo XCorr de 3.685559.

5. Conclusiones

Las modificaciones realizadas sobre el algoritmo de Comet-PTM, para incluir las Δ masas teóricas próximas en el cálculo del XCorr, logran mejorar la calidad de las identificaciones en aproximadamente un tercio de los péptidos (34.01%) analizados en este estudio.

Un valor de tolerancia de ± 1 Da es suficiente para hallar la Δ masa teórica correspondiente en aproximadamente un 89% de los casos, mientras que ampliar dicha tolerancia hasta ± 2 Da permite mejorar un 11% de dichas identificaciones, siendo indicado para el análisis de experimentos de tipo DIA, en los cuales el error es potencialmente mayor.

Se logra, además, reducir el ruido de fondo en el conjunto de Δ masas detectadas, de forma que en lugar de observar un conjunto de Δ masas repartidas en torno al valor teórico para una modificación dada, se consigue concentrar estos valores en un pico que corresponde exactamente con la Δ masa teórica, manteniendo separados únicamente aquellos valores de Δ masa experimental cuyo XCorr no mejora al ser sustituidos por el valor teórico, y que podrían corresponder a modificaciones no recogidas en la base de datos empleada.

Al determinar el valor exacto para la Δ masa de la PTM presente en un péptido, obtenemos resultados más relevantes, ya que se le puede asociar un significado biológico de forma más fiable. Además, el nuevo algoritmo permitiría utilizar una lista de Δ masas teóricas establecida por el usuario, permitiendo no sólo considerar las modificaciones descritas en UNIMOD, sino también emplear listas curadas de Δ masa según el conocimiento obtenido en el laboratorio.

Los incrementos de XCorr observados se justifican mediante el control de calidad con Vseq, que muestra cómo al sustituir la Δ masa experimental por la Δ masa teórica se completan las series de fragmentación, incluyendo fragmentos que anteriormente se estaban descartando debido a errores experimentales o saltos monoisotópicos. Además, el error de los fragmentos disminuye, de modo que se incluyen en el cálculo de la puntuación, lo que explica perfectamente el incremento del XCorr de los scans analizados.

RECOM representa un salto cualitativo en los análisis de modificaciones post-traduccionales libres de hipótesis mediante técnicas de proteómica por espectrometría de masas. Permite un análisis global sin precedentes de los mecanismos moleculares dependientes de PTMs que subyacen a diversos procesos fisiopatológicos. Asimismo, constituye una herramienta muy valiosa para el descubrimiento de nuevos biomarcadores asociados a estas enfermedades, con su consiguiente utilidad en la clínica.

5.1. Objetivos futuros

En este estudio se ha optado por trabajar con los datos crudos, sin filtrar por FDR, con el objetivo de aumentar el poder estadístico del análisis de los resultados al considerar un mayor número de puntos. El siguiente paso será calcular la FDR (False Discovery Rate) para validar las identificaciones. En nuestra experiencia, un filtrado al 5% FDR se traduciría aproximadamente en la eliminación de valores de XCorr inferiores a 1.8. Interesantemente, los mayores incrementos de XCorr observados se observan principalmente para valores de XCorr por encima de este umbral, y no se verían afectados por el filtrado.

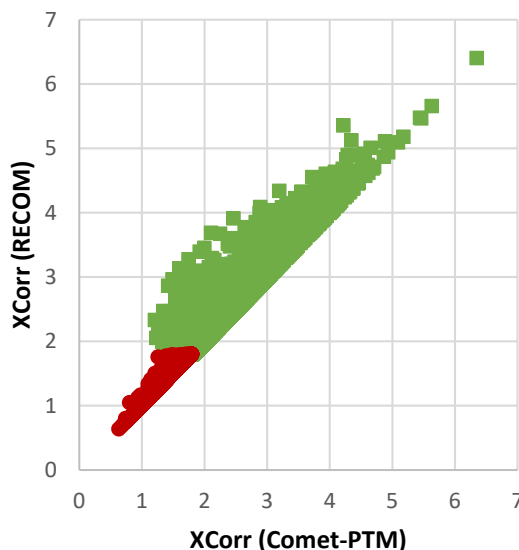


Figura 17. Representación aproximada de los resultados esperados si se aplicase un filtrado 5% FDR. La población de puntos rojos sería descartada.

Al analizar los resultados de este estudio se ha comprobado que el mejor candidato identificado para cada scan al aplicar RECOM continúa siendo el mismo (la secuencia peptídica no varía) tras la sustitución de la Δ masa experimental por la Δ masa teórica para todos los candidatos. Únicamente se observó un cambio de secuencia de los candidatos en 2 scans de los 10 000 analizados. En ambos casos se trataba de una diferencia mínima de XCorr (0.014 en el primer caso y 0.0274 en el segundo). Se trataba, además, de valores de XCorr inferiores a 1.8, que en nuestra experiencia no hubieran superado el corte de FDR. Esto indica que se podría acelerar en gran medida el proceso de búsqueda si, en lugar de calcular el XCorr teórico para todos los candidatos (Comet-PTM considera por defecto un máximo de 500 candidatos para cada scan), se realiza el cálculo únicamente para el primer candidato. Además, podría establecerse un criterio para el cual únicamente si los primeros n candidatos poseen valores de XCorr muy cercanos entre sí, se calculen los XCorr teóricos para cada uno de ellos, y se seleccione el nuevo mejor candidato.

Las modificaciones introducidas en Comet-PTM nos permiten considerar un conjunto de Δ masas teóricas que corresponden a modificaciones únicas. Una nueva mejora para este algoritmo sería añadir la capacidad de considerar combinaciones de modificaciones a partir de las masas correspondientes a modificaciones individuales. A partir de la información presente en los fragmentos, podría ser posible determinar la posición concreta dentro del péptido de cada modificación que forme parte de la combinación. De esta forma, podría conseguirse asignar a valores teóricos una proporción aún mayor de Δ masas experimentales.

Por último, se plantea también la posibilidad de introducir un método para corregir las puntuaciones de Xcorr. La fórmula empleada actualmente para el cálculo de XCorr (dependiente del número de fragmentos observados y la carga) da preferencia a péptidos de mayor tamaño, ya que frecuentemente presentan más fragmentos y cargas más altas que péptidos más cortos. Empleando un método para el cálculo de XCorr que puntúe péptidos incluyendo una corrección para la carga y la longitud ^[17], se daría prioridad al péptido más corto capaz de explicar los fragmentos observados.

6. Bibliografía

- [1] M. Mann, O.N. Jensen. **Proteomic analysis of post-translational modifications.** *Nat Biotechnol.*, 21 (2003), pp. 255-261. doi: [10.1038/nbt0303-255](https://doi.org/10.1038/nbt0303-255)
- [2] N. Bagwan, E. Bonzon-Kulichenko, E. Calvo, A.V. Lechuga-Vieco, S. Michalakopoulos, M. Trevisan-Herraz, I. Ezkurdia, J.M. Rodríguez, R. Magni, A. Latorre-Pellicer, J.A. Enríquez, J. Vázquez. **Comprehensive Quantification of the Modified Proteome Reveals Oxidative Heart Damage in Mitochondrial Heteroplasmy.** *Cell Reports*, 23 (2018), pp. 3685-3697. doi: [10.1016/j.celrep.2018.05.080](https://doi.org/10.1016/j.celrep.2018.05.080)
- [3] C. Delporte, K.Z. Boudjeltia, C. Noyon, P.G. Furtmülle, V. Nuyens, M.C. Slomianny, P. Madhoun, J.M. Desmet, P. Raynal, D. Dufour, C.N. Koyani, F. Reyé, A. Rousseau, M. Vanhaeverbeek, J. Ducobu, J.C. Michalski, J. Nève, L. Vanhamme, C. Obinger, E. Malle, P. Van Antwerpen. **Impact of myeloperoxidase-LDL interactions on enzyme activity and subsequent posttranslational oxidative modifications of apoB-100.** *J Lipid Res.*, 55 (2014), pp. 747-57. doi: [10.1194/jlr.M047449](https://doi.org/10.1194/jlr.M047449)
- [4] J. Cubedo, T. Padró, X. García-Moll, X. Pintó, J. Cinca, L. Badimon. **Proteomic signature of Apolipoprotein J in the early phase of new-onset myocardial infarction.** *J Proteome Res.*, 10 (2011), pp. 211-20. doi: [10.1021/pr100805h](https://doi.org/10.1021/pr100805h)
- [5] P. Natarajan et al. **Multiethnic Exome-Wide Association Study of Subclinical Atherosclerosis.** *Circ Cardiovasc Genet.* 9 (2016), pp. 511-520. doi: [10.1161/CIRCGENETICS.116.001572](https://doi.org/10.1161/CIRCGENETICS.116.001572)
- [6] M. Di Franco, B. Lucchino, F. Conti, G. Valesini, F.R. Spinelli. Asymmetric Dimethyl Arginine as a Biomarker of Atherosclerosis in Rheumatoid Arthritis. *Mediators Inflamm.* (2018), doi: [10.1155/2018/3897295](https://doi.org/10.1155/2018/3897295)
- [7] A.J. Link, J.K. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates. **Direct analysis of protein complexes using mass spectrometry.** *Nat. Biotechnol.*, 17 (1999), pp. 676-682. doi: [10.1038/10890](https://doi.org/10.1038/10890)
- [8] J. Griss, Y. Perez-Riverol, S. Lewis, D.L. Tabb, J.A. Dianes, N. Del-Toro, M. Rurik, M.W. Walzer, O. Kohlbacher, H. Hermjakob. **Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets.** *Nat. Methods*, 13 (2016), pp. 651-656. doi: [10.1038/nmeth.3902](https://doi.org/10.1038/nmeth.3902)
- [9] O.S. Skinner, N.L. Kelleher. **Illuminating the dark matter of shotgun proteomics.** *Nat. Biotechnol.*, 33 (2015), pp. 717-718. doi: [10.1038/nbt.3287](https://doi.org/10.1038/nbt.3287)
- [10] J.M. Chick, D. Kolippakkam, D.P. Nusinow, B. Zhai, R. Rad, E.L. Huttlin, S.P. Gygi. **A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides.** *Nat Biotechnol.* 2015 Jul;33(7):743-9. doi: [10.1038/nbt.3267](https://doi.org/10.1038/nbt.3267).
- [11] J.K. Eng, T.A. Jahan, M.R. Hoopman. **Comet: an open source tandem mass spectrometry sequence database search tool.** *Proteomics*, 1 (2013), pp. 22-24. doi: [10.1002/pmic.201200439](https://doi.org/10.1002/pmic.201200439).
- [12] J.K. Eng, A.L. McCormack, J.R. Yates. **An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database.** *J. Am. Soc. Mass Spectrom.* 5 (1994), pp. 976-989. doi: [10.1016/1044-0305\(94\)80016-2](https://doi.org/10.1016/1044-0305(94)80016-2)

- [13] D.M. Creasy, J.S. Cottrell. (2004), **Unimod: Protein modifications for mass spectrometry**. *Proteomics*, 4 (2004), pp. 1534-1536. doi: [10.1002/pmic.200300744](https://doi.org/10.1002/pmic.200300744)
- [14] Comet-PTM release 2018.01, <http://comet-ms.sourceforge.net/>, last revised 2018/12/05
- [15] UniProt release 2018_07, <https://www.uniprot.org/news/2018/07/18/release>, published July 18, 2018
- [16] Vseq, <https://github.com/CNIC-Proteomics/Vseq>
- [17] S. Martínez-Bartolomé, P. Navarro, F. Martín-Maroto, D. López-Ferrer, A. Ramos-Fernández, M. Villar, J.P. García-Ruiz, J. Vázquez. **Properties of Average Score Distributions of SEQUEST The Probability Ratio Method**. *Mol Cell Proteomics*. 7 (2008), pp. 1135-1145. doi: [10.1074/mcp.M700239-MCP200](https://doi.org/10.1074/mcp.M700239-MCP200)
- [18] S. Cogliati, E. Calvo, M. Loureiro, A.M. Guaras, R. Nieto-Arellano, C. Garcia-Poyatos, I. Ezkurdia, N. Mercader, J. Vázquez, J.A. Enriquez. **Mechanism of super-assembly of respiratory complexes III and IV**. *Nature* 539 (2016), pp. 579-582. doi: [/10.1038/nature20157](https://doi.org/10.1038/nature20157)